Procedures for Calibrating Hourly Simulation Models to Measured Building Energy and Environmental Data

J. S. Haberl
Associate Professor, Department of Architecture, Energy Systems Laboratory, Texas A&M University, College Station, TX 77843-3581 jhaberl@esl.tamu.edu.

T. E. Bou-Saada
Currently at Houston Light and Power, Houston, TX. Formerly a Research Associate at the Energy Systems Laboratory, Texas A&M University

This paper discusses procedures for creating calibrated building energy simulation programs. It begins with reviews of the calibration techniques that have been reported in the previous literature and presents new hourly calibration methods including a temperature bin analysis to improve hourly x–y scatter plots, a 24-hour weather-daytype bin analysis to allow for the evaluation of hourly temperature and schedule dependent comparisons, and a 52-week bin analysis to facilitate the evaluation of long-term trends. In addition, architectural rendering is reviewed as a means of verifying the dimensions of the building envelope and external shading placement as seen by the simulation program. Several statistical methods are also presented that provide goodness-of-fit indicators, including percent difference calculations, mean bias error (MBE), and the coefficient of variation of the root mean squared error (CV(RMSE)). The procedures are applied to a case study building located in Washington, D.C. where nine months of hourly whole-building electricity data and site-specific weather data were measured and used with the DOE-2.1D building energy simulation program to test the new techniques. Simulations that used the new calibration procedures were able to produce an hourly MBE of −0.7% and a CV(RMSE) of 23.1% which compare favorably with the most accurate hourly neural network models (Kreider and Haberl, 1994a, b).

Introduction

Computers and programmable calculators have been used extensively during the past three decades as effective heating, ventilating, and air-conditioning (HVAC) design tools to supplement tedious, manual energy calculations. Initially, super mini-computers or mainframe computers were required to perform the hourly simulations. During these early periods simulation was restricted to research organizations supported by public funding, large consulting firms and manufacturers, and a relatively few academic research institutions because of the large expenses incurred for computer hardware, and the limited availability of both software and qualified computer personnel (Ayers and Stamper, 1995). As computing technology has become affordable, engineers and architects have taken advantage of hourly simulation programs on desktop personal computers (PCs) that can inexpensively and quickly perform load calculations (ASHRAE, 1991). Eventually, simulation packages began to be used for retrofit evaluation in existing buildings which required calibration to measured data.

The calibration of a simulation to measured monthly utility data has been the preferred method for many years (Diamond and Hunn, 1981; Haberl and Claridge, 1985; McLain et al., 1993). Recently, studies have reported calibrated models using hourly measured data (Hsieh, 1988; Hinchey, 1991; Bronson, 1992; Kaplan et al., 1990; Clarke et al., 1993; Manke and Hittle, 1996). Most of the previous methods have relied on very simple comparisons including bar charts, monthly percent difference time-series graphs, and monthly *x-y* scatter plots. Unfortunately, at hourly levels of calibration, many of the traditional graphical calibration techniques become overwhelmed with too many data points which makes it difficult to determine the central tendency of the black clouds of data points. A few advanced methods have been proposed including carpet plots (Bronson, 1992), and comparative 3-D time-series plots (Hinchey, 1991).

Calibrating computer models to actual metered data is not a new practice. As early as 1970, recommendations were made to calibrate models based on measured data (Ayers and Stamper, 1995). Some researchers and engineers have attempted to compile "how to" manuals and methods in order to simplify this task; however, in almost all cases the end result typically falls short of a complete procedure (Hsieh, 1988; Kaplan et al., 1992; Bronson et al., 1992). To date, no consensus standards have been published on calibration procedures and methods that can generically be used on a wide variety of buildings. Historically, actual calibration has been an art form that inevitably relies heavily on user knowledge, past experience, statistical expertise, engineering judgment, and an abundance of trial and error.

One problem with reporting simulation accuracy rests with the calculation procedures which have been reported in the previous work. Typically, when a model is established as being calibrated (i.e., the user reports that the accuracy for electricity is ±5% per month), the author does not reveal the techniques used other than stating the final result. Hourly or daily error values are seldom reported. Even in cases when error estimates are presented, the methods and equations used to obtain the comparisons are not. A complete review of the methods is provided in Bou-Saada (1994a, b) and Bou-Saada and Haberl (1995a, b).

Methodology

To simulate and calibrate a computer model to measured energy data, several stages needed to be accomplished. First, site specific hourly weather data included dry bulb temperature, relative humidity, and peak wind speed gathered from the nearby National Weather Service (NWS) station at the Wash-

Contributed by the Solar Energy Division of The American Society of Mechanical Engineers for publication in the ASME JOURNAL OF SOLAR ENERGY ENGINEERING. Manuscript received by the ASME Solar Energy Division, May 1998; final revision, May 1998. Associate Technical Editor: D. E. Claridge

ington National Airport. Global solar radiation data was also measured on-site. A routine developed by Erbs et al. (1982) was then used to convert global solar radiation measured on an 18 deg south facing tilt into global horizontal beam and diffuse radiation (Bou-Saada, 1994a, b). The weather data were then joined into a single datafile and packed onto a TRY weather tape (Bronson, 1992) for use with the DOE-2 simulation program. Next, energy use and solar data from two dataloggers were merged with the NWS data for use with the graphical techniques which will be described shortly.

The calibration procedure also entailed creating a DOE-2 input file based on information obtained from site visits and architectural as-built plans. The hourly output report from the simulation was processed further with post-processing routines (Bronson, 1992) specially modified for this work. This avoids the long-term data averaging encountered with monthly simulation comparisons (Hinchey, 1991).

Information Required for Calibrating DOE-2. Figure 1, adapted from Bronson (1992), is a general overview of a calibration process (Bou-Saada, 1994a). Figure 2 illustrates the specific data and graphical processing that is involved in the calibrated procedures used in this paper. The grouping at the top of Fig. 1 includes all the required input information to produce a DOE-2 simulation including: DOE-2 reference manuals, as-built drawings, information from site visits, utility billing data, on-site measured energy consumption data and on-site measured weather data. A typical DOE-2 input file may be produced using any number of computerized ASCII text editors. Any DOE-2 simulation usually requires a visit to the standard set of DOE-2 reference manuals to observe correct BDL syntax (the DOE-2 input format) and mandatory BDL requirements (LBL, 1980, 1981, 1982, 1989).

As-built drawings help to correctly dimension the building and calculate lighting and equipment levels. A site visit is generally essential to verify lighting fixture counts and equipment nameplate data as well as to verify dimensions and any other miscellaneous discrepancies. The site visit should also include photographs of the building's surroundings for establishing shading calculations and detailed interviews with occupants, engineers, architects, and building operations personnel. Also included in the site visit should be shading measurements, and clamp-on RMS Wattage measurements of key pieces of equipment to verify actual power requirements.

A major part of the site visit includes the gathering of energy use data and/or monthly utility bills. Either is acceptable, but neither is a strict requirement to compile a first pass at the input file. An HVAC system air balance report is also helpful when describing the zone air flow rates. On-site weather data measured for the simulation period have also been shown by Haberl et al. (1995) to significantly improve the simulation. In cases where no weather data are available, standard weather tapes such as TRY and TMY may be used. Finally, prior experience with the DOE-2 program plays a crucial factor that can benefit the user in avoiding commonly made mistakes. Many problems with the input file may be avoided simply by having prior knowledge of program expectations as well as a thorough engineering understanding of HVAC systems and building heat loss/gain in general.

The flowchart provided in Fig. 2 illustrates the large number of software modules that were required to collect the measured data from the site, check the data quality, archive the data, convert various site measurements into meaningful units[1] for use by DOE-2, pack the weather tape, run DOE-2, extract the hourly data from DOE-2 and merge it with measured data, prepare the hourly data for graphical processing, and produce the graphical comparison plots.

After a simulation is performed, the Statistical Analysis Software (SAS, 1989) program is used to analyze the goodness-of-fit and produce graphical feedback of the simulation progress. This includes time-series plots, bin plots, and three-dimensional hourly plots for further analysis. This allows for a graphical comparison of the simulated consumption to the monitored consumption, and statistical comparison of the simulated and measured consumption. With this information now processed, the user can then decide if the model is calibrated to an acceptable level, and of equal importance, where the remaining mismatch may be located. This second feature is accomplished with the assistance of the calibration tools. If it is determined that a simulation is not fully calibrated, the areas where the simulated data do not match the measured data must be identified and adjusted in the input file. The DOE-2 program is run again and the data processed for comparison until an acceptable calibration is reached. The following sections describe the specific calibration methods that were found to be helpful.

Architectural Rendering. Several software programs have recently become available for purposes of architectural rendering or viewing of building simulation input files (Degelman, 1995; Hirsch et al., 1995). One such program, DrawBDL (Huang, 1993), was used to verify the building envelope descriptions used in the DOE-2 input file. The building shown in Fig. 3 shows a view of the case study building using the DrawBDL program. The software also includes such capabilities as rotating the building to allow for viewing from any direction and includes views of a three-dimensional perspective, a plan view, an elevation view, and a wire frame view. With a BDL visualization tool, each case study building envelope surface and shading surface can be inspected for proper placement, size, and orientation. This type of checking could not easily be performed prior to the creation of such architectural rendering tools.

Graphical Methods for Improving a Calibration. Traditionally, DOE-2 users have used simple time-series plots of simulated and actual data which are superimposed upon the same graph for a short period to confirm their calibrations, for example in Hsieh (1988), Hunn et al. (1992), and Reddy (1993). Although two-dimensional time-series plots are useful for determining certain features, a special problem exists when plotting long-term hourly time-series data. In such cases, direct comparison becomes ineffective for all practical purposes because it is very difficult to identify individual hourly data points.

One improvement over past graphical techniques is shown in Fig. 4 which shows an example of a binned analysis that was modified for this paper from indices developed by Abbas (1993). The superimposed and juxtaposed binned box-whisker-mean plots display the maximum, minimum, mean, median, 10^{th}, 25^{th}, 75^{th}, and 90^{th} percentile points for each data bin for a given period of data. These plots eliminate data overlap and allow for a statistical characterization of the dense cloud of hourly points (scatter plots are still useful in showing individual point locations). The important feature to note about this plot is that the data are statistically binned by temperature. This feature allows for the bin-by-bin goodness-of-fit to be evaluated quantitatively. Using the box-whisker-mean plot combined with a scatter plot also allows one to visualize the data as a whole while simultaneously seeing the effects of the outliers in specific situations. Both of these features are important to the efficiency with which the graph conveys an accurate and consistent message to the viewer (Tukey, 1977; Cleveland, 1985).

In Fig. 4 only the weekday data are plotted, using a technique developed by Abbas (1993) and modified for this paper, that includes a combination of vertical and horizontal juxtapositioning, temperature-based box-whisker-mean bins and super-positioning of the measured mean bin line in the

[1] The primary need for this was in the conversion of global solar radiation to diffuse and beam radiation. This was accomplished with routines developed by Erbs et al. (1982). Additional processing involved the analysis of electricity produced by the photovoltaic routines, and conversion of weather measurements into units acceptable to the DOE-2 weather packer.

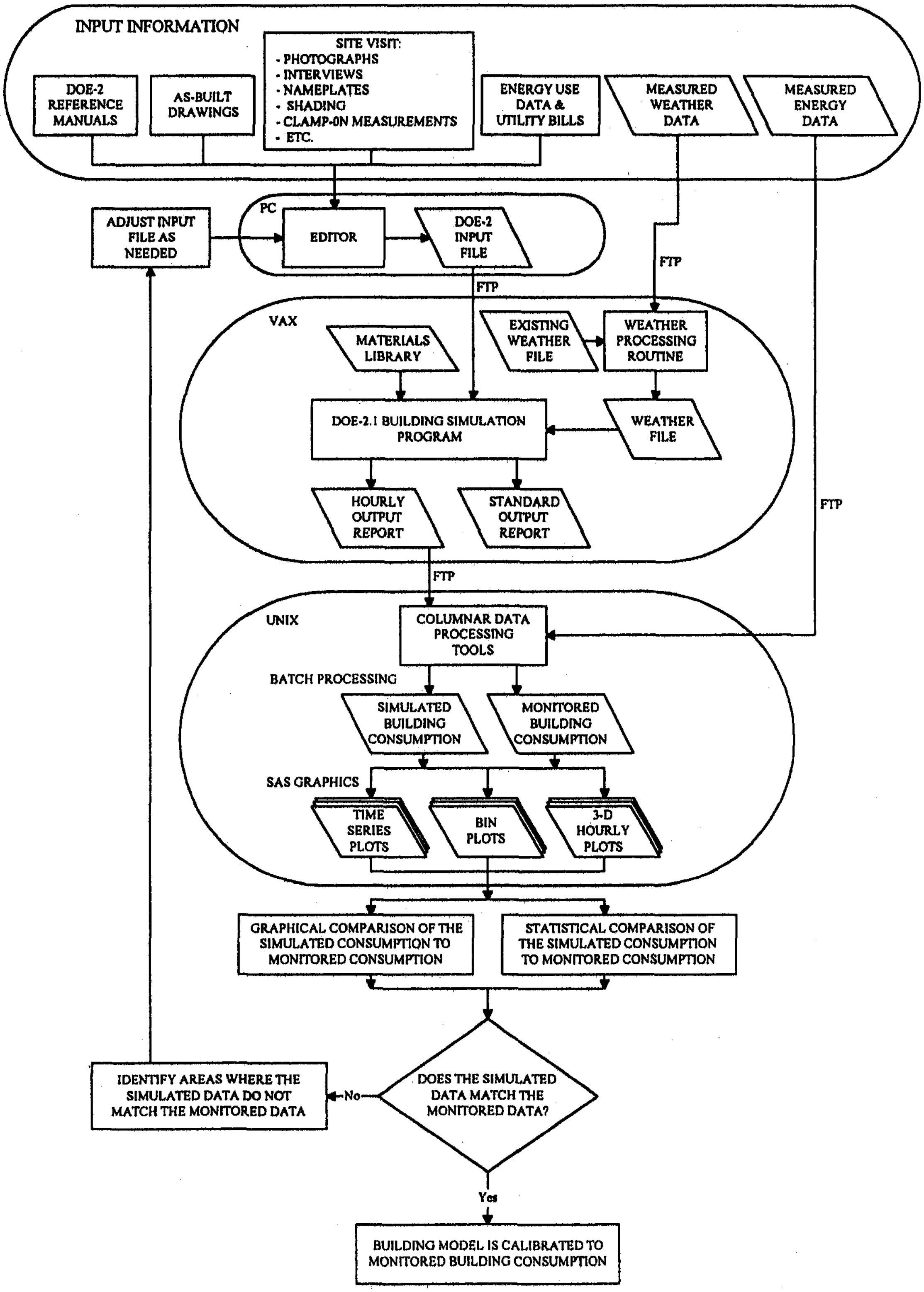

Fig. 1 DOE-2 calibration procedure. This graph illustrates the information flow of the calibration process beginning with the input information, editing, simulation, graphical post-processing, and comparison against the measured data

lower right graph. Similar analysis can be performed with weekend and holiday data (Bou-Saada, 1994a). In the upper left graph the hourly measured whole-building electricity use is plotted against hourly ambient temperature. In the upper right graph, the corresponding DOE-2 simulated electricity data for the same period are shown. Below each scatter plot are binned box-whisker-mean plots. These plots show the whole-building electricity consumption as a function of outdoor temperature bins divided into 10°F segments. One final feature of these plots is that the measured data mean is superimposed as a dashed line onto the calibrated DOE-2 simulation data. The variation in mean lines in each bin provides a measure of how well the model is calibrated at a specific temperature bin. Likewise, the inter-quartile range (i.e., the

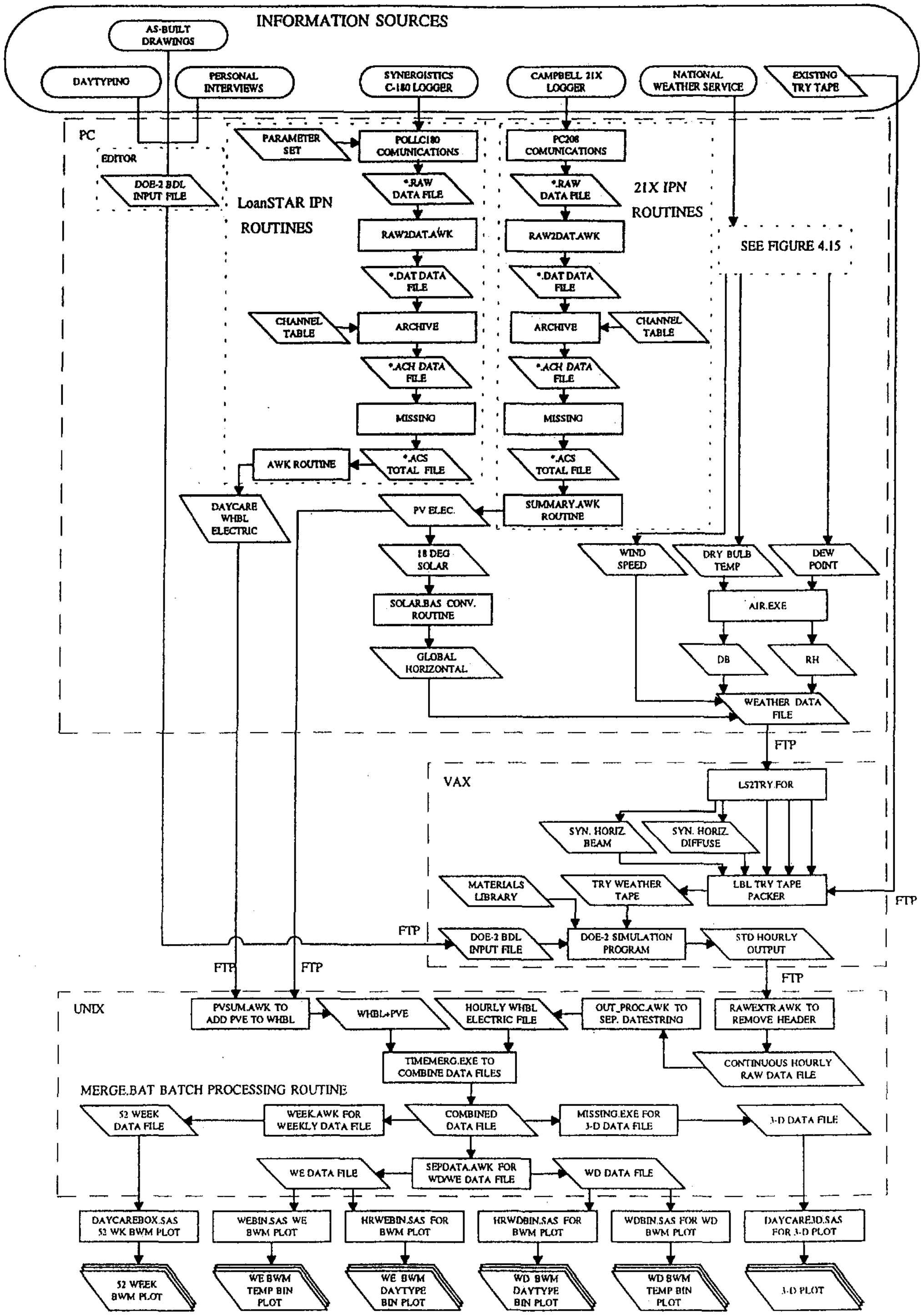

Fig. 2 Data and graphical processing flowchart. This figure shows the sequence of processing routines that were used to develop the graphical calibrations presented in this paper (Bou-Saada, 1994a, b)

distance between the 25th and 75th percentiles) represents the hourly variation in a given bin.

Figure 5 is an example of a weekday 24-hour weather daytype box-whisker-mean plot that shows the whole-building electricity use versus the hour-of-the-day for both the measured data and the DOE-2 simulated data in three weather daytypes. A similar analysis can also be performed with weekend and holiday data. The weather daytypes arbitrarily[2] divide the measured data into bin temperatures below 45°F, between 45°F and 75°F, and above 75°F. The original concept for this plot can be traced to the weather daytype analysis developed by Hadley (1993).

This additional calibration procedure allows a DOE-2 user to view and analyze the weather and schedule dependent hourly energy use. The solid line in parts (b), (d), and (f) is the simulated mean. The dashed line is the measured

[2] The 45°F and 75°F separation points were chosen arbitrarily for this particular building. A more formal routine could be used as well. For example, an iterative search routine could be used to determine the optimum separation points by choosing those points that yield the lowest average inquartile range across all bins for all days. Another example might be to analyze the daily weekday-weekend data with the 5 parameter Princeton Scorekeeping Method (PRISM) Fels (1986).

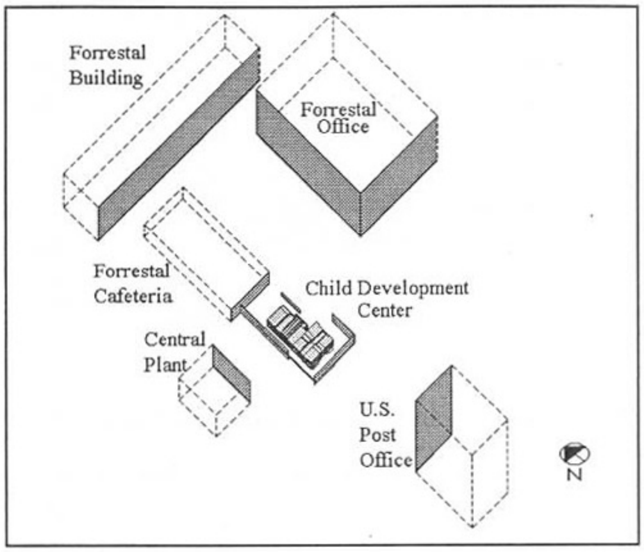

Fig. 3 The U.S.D.O.E. Forrestal complex and surrounding area. The solid planes represent shading from buildings, walls, and trees

mean line from parts (a), (c), and (e) that is superimposed onto the simulated data. These plots confirmed that the building's 24-hour electricity profiles are strongly influenced by the ambient temperature. The plots also provide a more efficient method of viewing the data based on heating only, no heating or cooling, and cooling only modes.[3]

Comparative 3-D Surface Calibration Plots. In Figs. 6(a)–6(d) comparative 3-D surface plots show the monitored data in part (a) and the DOE-2 simulated data in part (b) for the April to December 1993 period. Figure 6(c) shows positive-only values of the measured data subtracted from the DOE-2 predicted data while Fig. 6(d) shows positive only values of the DOE-2 predicted data subtracted from the measured data.

Individual hourly differences may be visually detected over the entire simulation period using these plots which allows the user to recognize patterns in the comparisons. For example, DOE-2's over-predictions in the spring and fall mornings and afternoons (Fig. 6(c)) and both over- and under-predictions in the late evening throughout the year (Fig. 6(c) and 6(d)). An obvious benefit of such plots is their ability to aid in the identification of oversights such as a daylight savings shift or misalignment of 24-hour holiday profiles (Bronson et al., 1992). One negative drawback associated with these graphs is the difficulty in viewing exact details such as the specific hour, magnitude or specific day on which a misalignment occurs.

By inspecting Fig. 6(d), it is clear that DOE-2 is under-predicting early morning unoccupied electricity use in the spring between April 1 and July 1 1993. One possible cause for this is the continuous operation of the building's cross-wired exhaust fans that were supposed to be turned-off,[4] but operated continuously. This draws in cool outdoor air in the spring and can activate the heating system when the temperature falls below the setback temperature of 55°F. DOE-2 will not simulate this condition properly because the program cannot account for negative building pressures. Therefore, no extra heating is evident in the simulated plot. It is felt that this also causes the under-prediction of energy use in the early morning hours of April and December as shown in Fig. 6(d). The ragged on/off "pickets" during the morning hours of July through September are a characteristic of DOE-2's hourly calculation algorithm when systems are scheduled normally off with the exception of a minimum zone temperature.

Figure 7 shows the same data as the 3-D surface plots in Fig. 6(a) and 6(b). However, Fig. 7 displays the energy usage using weekly time-series box-whisker-mean plots. The measured data are shown in part (a) and the DOE-2 simulation is shown in part (b). The x-axis in Fig. 7 is the simulation week number, for this simulation, week "0" begins on April 1. The y-axis shows the whole-building electricity use in both (a) and (b).

Figure 7 utilizes graphical superposition of the mean line from Fig. 7(a) (dashed line) upon Fig. 7(b) to further improve the viewing effectiveness of the graph. It follows a markedly similar path traced by the simulated data mean line represented by the solid line. Fine differences can also be seen in those points above the 90th percentile in weeks 0–5. Those points represent the hours of electrical resistance heating in the measured data that occurred when the staff manually switched on the baseboards in the toddler rooms to preheat the rooms prior to the arrival of the children.

When evaluating each set of paired weekly bins, it may be concluded that the average weekly simulated data seems to consistently track the average weekly measured data since the simulated box sizes (i.e., inter-quartile range) do not deviate significantly from the measured box sizes. However, the minimum simulated data limits are consistently lower in the DOE-2 plot (Fig. 7(b)) further emphasizing the difficulties encountered in predicting the nighttime building shut down schedule. The unusual consistency of the 25th percentile and minimum points seen in the winter is a characteristic of the DOE-2's rigid scheduling. It would appear that beginning in approximately week 11 and ending in week 21, the measured nighttime HVAC setback mode was overridden due to the zone temperature exceeding the control system upper setpoint temperature limit. This can be seen by the rise in the median for each week, beginning in week 11 and ending in week 21.

Statistical Calibration Methods. In the previous research very little has been said about the quantitative methods that were used to declare the model "calibrated." However, some guidance has been provided in a few papers. In Haberl et al. (1995) and Bronson (1992) the simulation results were summed to monthly values and verified against the actual measured data via a percent difference calculation. Torres-Nunci (1989) and Hinchey (1991) declared their models "calibrated" and submitted hourly graphs to demonstrate the goodness-of-fit. They also provided numerical differences only in the form of ± monthly differences. The problem with this approach is that the ± monthly difference calculation does not provide a fine enough goodness-of-fit indicator and is in fact misleading because it can indicate a near perfect fit when there is still considerable hourly error in the calibration. Therefore, in the interest of furthering the calibration procedures, several statistical calculations were compared including a monthly mean difference, an hourly mean bias error (MBE) for each month, an hourly root mean squared error (RMSE) reported monthly, and an hourly coefficient of variation of the root mean squared error (CV(RMSE)) (Kreider and Haberl, 1994a, b; Haberl and Thamilseran, 1996). These indices were previously shown to be useful in comparing hourly neural network models against measured hourly use.

The percent difference is a simple calculation whereby a difference for each monthly measured and simulated energy use total is taken and divided by the measured monthly total consumption. This index is the typical value reported for most DOE-2 predictions (Diamond and Hunn, 1981; Kaplan et al., 1990, 1992; Bronson, 1992; McLain et al., 1993; Haberl et al., 1995). The mean bias error, MBE (%) (Kreider and Haberl, 1994a, b; Haberl and Thamilseran, 1996), is a method with which to determine a non-dimensional bias measure (the sum

[3] The plots may not be as useful for buildings that have simultaneous heating and cooling.

[4] According to the as-built drawings.

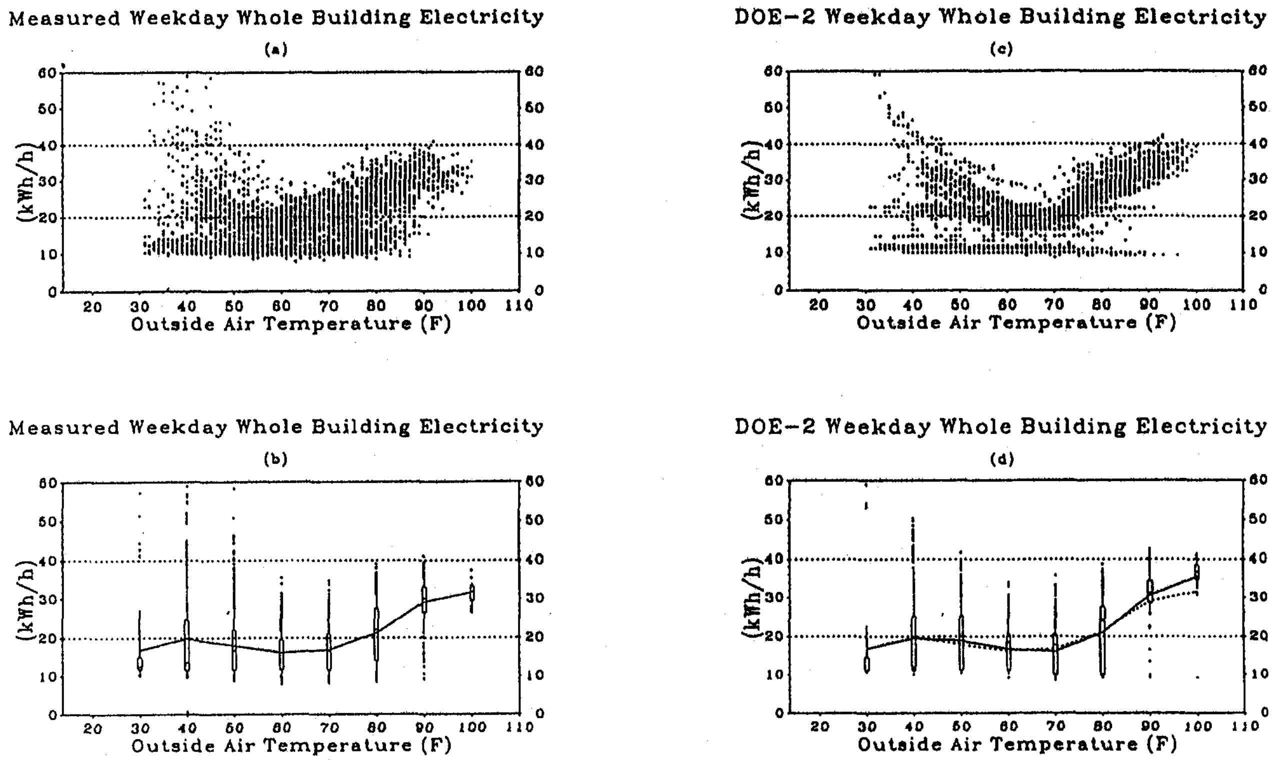

Fig. 4 Weekday temperature bin calibration plots. This figure shows the measured and simulated hourly weekday data as scatter plots against temperature in the upper plots and as binned box-whisker-mean plots in the lower plots

of errors) between the simulated data and the measured data for each individual hour (Katipamula, 1994).

The root mean squared error (RMSE) is typically referred to as a measure of variability, or how much spread exists in the data. For every hour, the error, or difference in paired data points is calculated and squared. The sum of squares errors (SSE) are then added for each month and for the total periods and divided by their respective number of points yielding the

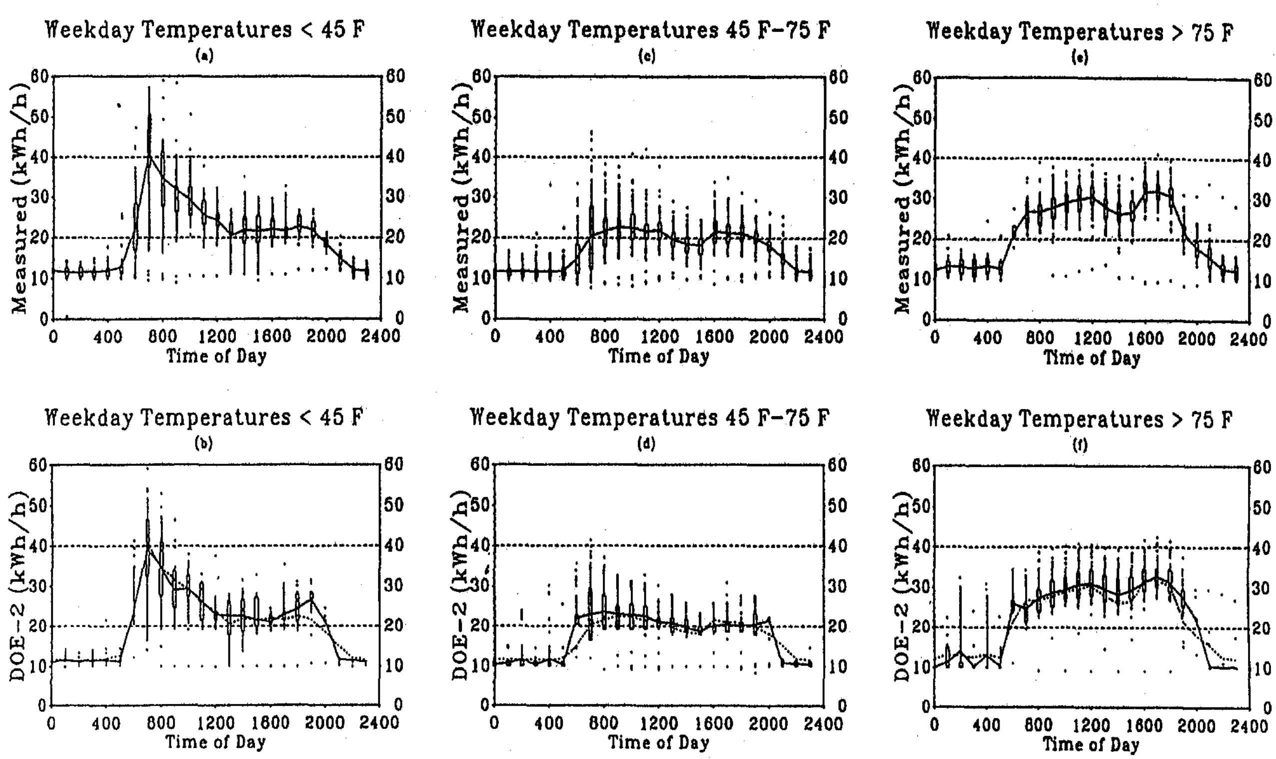

Fig. 5 Weekday 24-hour weather daytype box-whisker-mean plot. This figure shows the same information as shown in Fig. 4 displayed as weather daytype 24-hour box-whisker-mean plots for measured data (upper plots) and simulated data (lower plots)

mean squared error (MSE); whether for each month or the total period. A square root of the result is then reported as the root mean squared error (RMSE).

The coefficient of variation of the root mean squared error, CV(RMSE) (%) (Draper and Smith, 1981) is essentially the root mean squared error divided by the measured mean of all the data, a convenient way of reporting a non-dimensional result. CV(RMSE) allows one to determine how well a model fits the data; the lower the CV(RMSE), the better the calibration (i.e., the "model" in this case is the DOE-2 predicted data). Therefore, a CV(RMSE) is calculated for hourly data and presented on both a monthly summary and total data period. The equations for CV(RMSE) and MBE are included in the appendix of this paper.

The purpose of calculating the CV(RMSE) and comparing the results with the standard percent difference is to demonstrate that the use of a percent difference report only may be misleading. Since these calculations are usually shown for monthly simulations or even total simulation periods, the reader is never certain if the model is a true representation of the actual building or if the ± errors have canceled out. If one examines the hour-by-hour data results, it would be evident that each pair of points would in all likelihood be dissimilar and in some cases be significantly different, despite using the same measured weather data to drive the simulation model. Reporting monthly data therefore does not take into account the canceling out of individual differences observed when the simulation over-predicts during one hour and under-predicts during the next hour by approximately the same amount. More information on the statistical indices used in this paper can be found in Bou-Saada (1994a).

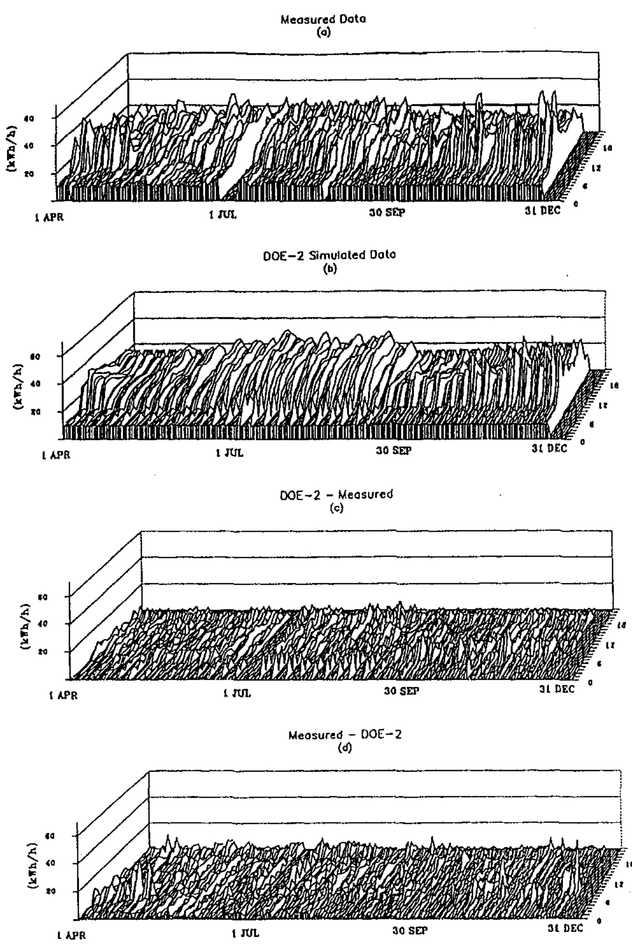

Fig. 6 Comparative three-dimensional plots. (*a*) measured data, (*b*) simulated data, (*c*) simulated-measured data (*d*) measured-simulated data

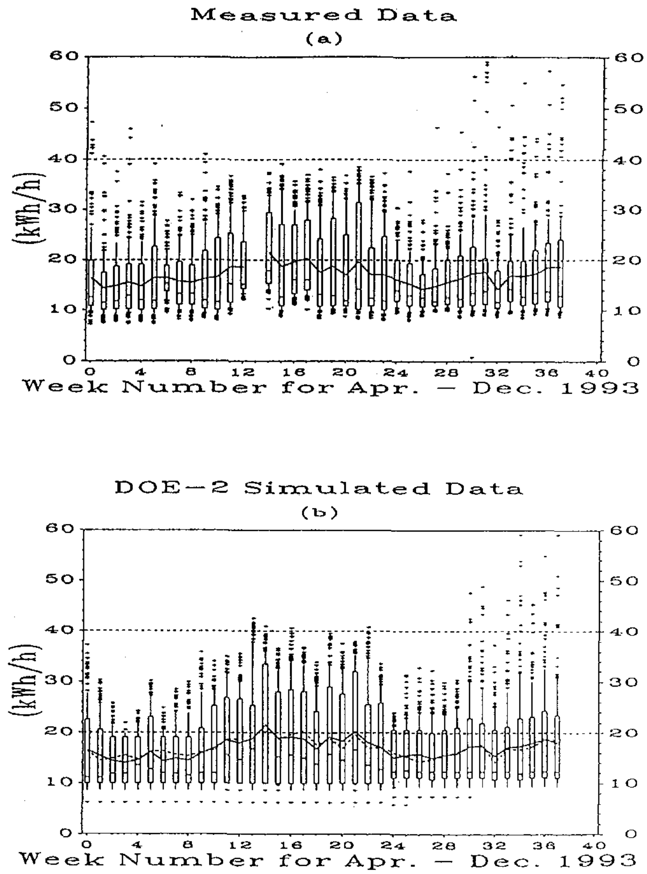

Fig. 7 52-week binned box-whisker-mean plots for the measured data (upper plot), and simulated data (lower plot)

The statistical results for the DOE-2 model can be seen in Tables 1(*a*) through 1(*d*). These tables show monthly and total measured kWh/month, measured hourly mean kWh, total simulated kWh/month, and simulated hourly mean kWh. More importantly, the table compares the monthly and total percent difference to the hourly MBE, hourly RMSE, and hourly CV(RMSE) for each month and the total period. The MBE, RMSE, and CV(RMSE) statistics are calculated for each hour and summed for each month and for the total period, the weekday occupied period, the weekday unoccupied period, and the weekend period.

Table 1*a* and Fig. 8(*a*) summarize the statistics for the total calibration period from April 1, 1993 through December 21, 1993. The contrast between percent difference and CV(RMSE) clearly can be seen when comparing May and July. In May, a simulated under-prediction exists to the tune of 4.9% while the CV(RMSE) is a relatively low 21.8% (i.e., low when compared to the other months in the table). July shows an under-prediction difference of 1.7% whereas the CV(RMSE) is a high 24.7%. The hourly peak and trough cancellation effect described earlier is evident here showing an overall low percent difference in July and an hourly high CV(RMSE). Despite all the months generally showing a low percent difference, the hourly CV(RMSE) consistently remained in the 20 to 27% range. This is still well within the range reported in a recent ASHRAE sponsored contest (Kreider and Haberl, 1994a, b). Clearly, calibrations should be performed using both hourly CV(RMSE) and hourly MBE calculations rather than solely with the total percent difference calculation.

In Tables 1*b* through 1*d* additional useful information can be seen as well. For example, in the weekday occupied period (Fig. 8(*b*) and Table 1*b*) lower CV(RMSE) can be seen for the air conditioning periods whereas higher CV(RMSE) pre-

Table 1 Statistics summary. 1(*a*) Total period, (1*b*) weekday occupied period, (1*c*) weekday unoccupied period statistics summary, (1*d*) weekend period. Values are given for the total and simulated monthly electricity use (kWh/mo), mean hourly electricity use for the month (kWh/h), and indicate the mean bias error (MBE), root mean squared error (RMSE), and coefficient of variation of the root mean square error (CV(RMSE)).

Total Period Statistics Summary

Month	Total Measured (kWh)	Mean (kWh)	Total Simulated (kWh)	Mean (kWh)	Total Difference %	Hourly MBE %	Hourly RMSE (kWh)	Hourly CV(RMSE) %
April	10,606	15.5	10,243	15.0	-3.4	-3.4	4.2	26.9
May	11,507	15.5	10,984	14.7	-4.9	-4.9	3.4	21.8
June	11,568	17.4	11,313	17.0	-2.2	-2.2	3.5	20.2
July	11,247	19.9	11,052	19.6	-1.7	-1.7	4.9	24.7
August	12,621	18.6	12,531	18.5	-0.7	-0.7	4.0	21.7
September	11,885	16.5	12,012	16.7	1.1	1.1	3.7	22.5
October	11,216	15.1	11,413	15.3	1.8	1.8	3.5	23.2
November	11,944	16.7	12,202	17.0	2.2	2.2	4.0	23.7
December	9,220	18.3	9,337	18.5	1.3	1.3	4.2	22.8
Total	101,814	16.8	101,051	16.8	-0.7	-0.7	3.9	23.1

Weekday Occupied Period Statistics Summary

Month	Total Measured (kWh)	Mean (kWh)	Total Simulated (kWh)	Mean (kWh)	Total Difference %	Hourly MBE %	Hourly RMSE (kWh)	Hourly CV(RMSE) %
April	5,233	21.3	5,172	21.0	-1.2	-1.2	5.6	26.1
May	5,379	21.3	5,334	21.2	-0.8	-0.8	3.4	15.9
June	6,127	25.6	6,169	25.8	0.7	0.7	3.6	13.9
July	5,859	29.3	6,246	31.2	6.6	6.6	4.1	13.9
August	6,997	29.9	6,886	29.4	-1.6	-1.6	3.6	11.9
September	6,246	23.7	6,438	24.4	3.1	3.1	4.5	19.1
October	5,292	21.0	5,440	21.6	2.8	2.8	5.1	24.1
November	6,153	23.4	6,410	24.4	4.2	4.2	5.2	22.4
December	4,932	27.4	5,149	28.6	4.4	4.4	6.2	22.6
Total	52,217	24.5	53,244	25.0	2.0	2.0	4.6	18.9

Weekly Unoccupied Period Statistics Summary

Month	Total Measured (kWh)	Mean (kWh)	Total Simulated (kWh)	Mean (kWh)	Total Difference %	Hourly MBE %	Hourly RMSE (kWh)	Hourly CV(RMSE) %
April	3,486	13.9	3,210	12.8	-7.9	-7.9	3.7	26.8
May	3,404	13.5	3,140	12.5	-7.8	-7.8	3.6	26.3
June	3,168	13.5	3,037	13.0	-4.1	-4.1	3.6	26.6
July	3,133	15.8	2,807	14.2	-10.4	-10.4	5.4	33.8
August	3,128	13.8	3,147	13.9	0.6	0.6	4.5	32.3
September	3,504	13.3	3,543	13.4	1.1	1.1	3.5	26.6
October	3,285	13.0	3,463	13.7	5.4	5.4	2.8	21.4
November	3,628	13.9	3,732	14.3	2.9	2.9	3.5	24.9
December	2,695	14.9	2,633	14.6	-2.3	-2.3	2.9	19.5
Total	29,430	13.9	28,712	13.5	-2.4	-2.4	3.8	27.0

Weekend Period Statistics Summary

Month	Total Measured (kWh)	Mean (kWh)	Total Simulated (kWh)	Mean (kWh)	Total Difference %	Hourly MBE %	Hourly RMSE (kWh)	Hourly CV(RMSE) %
April	1,887	10.0	1,860	9.8	-1.5	-1.5	2.0	20.1
May	2,725	11.4	2,474	10.3	-9.2	-9.2	3.2	27.9
June	2,272	11.8	2,106	11.0	-7.3	-7.3	3.4	28.5
July	2,255	13.5	2,000	12.0	-11.3	-11.3	5.3	39.2
August	2,496	11.5	2,497	11.5	0.0	0.0	4.1	35.4
September	2,135	11.1	2,031	10.6	-4.9	-4.9	2.6	23.2
October	2,639	11.0	2,511	10.5	-4.9	-4.9	1.7	15.0
November	2,164	11.2	2,060	10.7	-4.8	-4.8	2.1	18.9
December	1,593	11.1	1,556	10.8	-2.4	-2.4	1.6	14.3
Total	20,167	11.4	19,095	10.8	-5.3	-5.3	3.1	27.0

vailed in during the unoccupied periods. This is indicating that the use of the air conditioning was consistent during these periods during occupied periods. However, during unoccupied periods during the weekdays (Fig. 8(*c*) and Table 1*c*) the simulation did not match the measured data as well as these were periods of infrequent air-conditioning only when indoor condi-

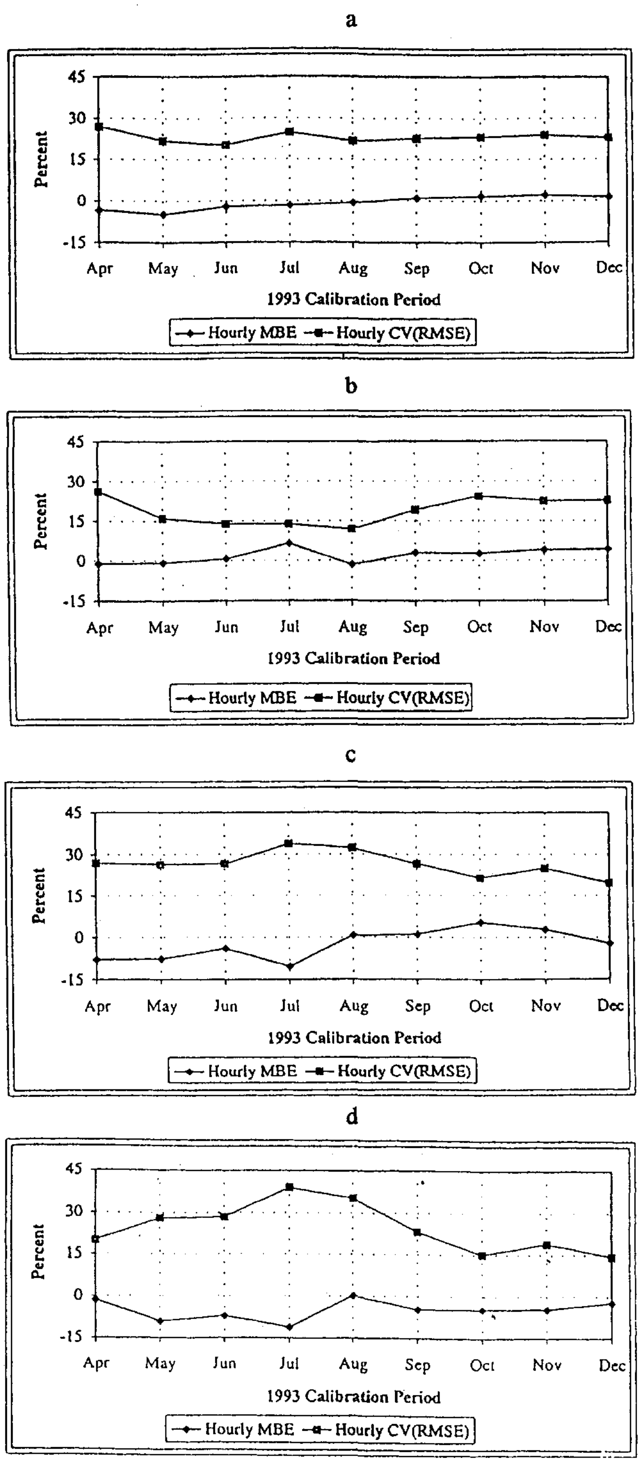

Fig. 8 Monthly calibration summaries. Graphical summaries of the information presented in Tables 1*a* through 1*d* are shown here for the total period (Fig. 8(*a*)), the weekday occupied period (Fig. 8(*b*)), the weekday unoccupied period (Fig. 8(*c*)), and the weekend period (Fig. 8(*d*))

tions exceeded a certain amount or systems were left running all night. During the weekends it was very difficult to predict the electricity use (Fig. 8(*d*) and Table 1*d*) because the systems were only run when indoor conditions exceeded a setpoint temperature, or when turned-on manually by staff.

Model Fine Tuning Progress. To represent the adequacy of the DOE-2 models from an initial workable model to a final calibrated model, each major change to the input file was documented through the research phase. A condensed set of iterations is graphically displayed in Fig. 9 which shows the impact of most of the major modifications made to the model. The corresponding history of the changes is listed in Table 2 which details the hourly total MBE and CV(RMSE) for each model and modifications. In reality, about 100 different iterations were run, however due to space constraints, a summary of each group of runs is shown in Fig. 9 and Table 2.

The first DOE-2 model to run without errors was considered Run #1 and is labeled as the base model. This run consisted of the best information that could be obtained from as-built drawings, equipment lists, audit findings, and DOE-2 default values. As Table 2 shows, the error is large with a 94.8% MBE and a 102.4% CV(RMSE). Run #2 showed slight improvements with the change of infiltration from 0.0 to 0.1 air changes per hour (ACH), extraction of hourly data from the PLANT sub-program instead of from the SYSTEMS sub-program, and modifications made to the indoor use of indoor lighting and equipment. Correcting the cooling and heating capacities to the manufacturer's specifications in Run #3 worsened the MBE and CV(RMSE) slightly calling for adjustments to the model elsewhere. The most significant modification to the model occurred with Run #4 when indoor lighting and equipment schedules were lowered to match the measured daily profiles. Clearly, one may conclude for this building that the lighting and equipment schedules had a significant impact on the simulation. This is consistent with advice given by Hsieh (1988), Kaplan et al. (1992), and Grifiths and Anderson (1994).

Run #5 was the first simulation to make use of the DrawBDL software to verify the envelope. As was the case with this simulation, corrections to wall locations were minor because they had been placed fairly close to where they belong in the first place. Run #5 also did not include the effects of shading at the site, thus not affecting the output significantly. Run #6 was another notable improvement in the process in which a large change in the MBE and CV(RMSE) was observed. This run incorporated a combination of adjustments to the ventilation air changes per hour, exhaust air volume, and power level corrections as specified by the balance report, adjustment of lighting power in one zone, addition of building shading, addition of specified ground temperatures, and the adjustment of cooling and heating schedules from 24-hour operation to "off at night" schedules. The cooling and heating load adjustment would in all probability be the most influential factor of this list by reducing the PLANT load significantly in the night setback mode.

As Fig. 9 and Table 2 show, the remainder of the runs were fine-tuning runs with no large scale change in either the MBE or CV(RMSE). Most of the changes made to the model included minor schedule adjustments until Run #12. By reviewing the input file, an error was detected in the original assumptions concerning the SYSTEMS input file. It was discovered that the incorrect air volume was originally specified by the inadvertent omission of the return air fan and volume. Correcting this error required a recalibration effort by adjusting the lighting and equipment schedules to compensate for the return air fan correction.

Run #12, the final run, was then considered the "calibrated" version. The MBE of −0.7% and an hourly CV(RMSE) of 23.1% were considered acceptable for the project. Previous work reported by Kreider and Haberl (1994a, b) showed that even the very best empirical models (i.e., artificial neural networks used with a large commercial building) were only capable of producing CV(RMSE) in the 10 to 20% range. When comparing the neural network results to the daycare center, which is a much smaller building, the CV(RMSE) of 23.1% is acceptable. For this paper, both CV(RMSE) and MBE did not appear to further improve without major modifications to the input file such as additional schedules to account for differences in the occupants' behavior.

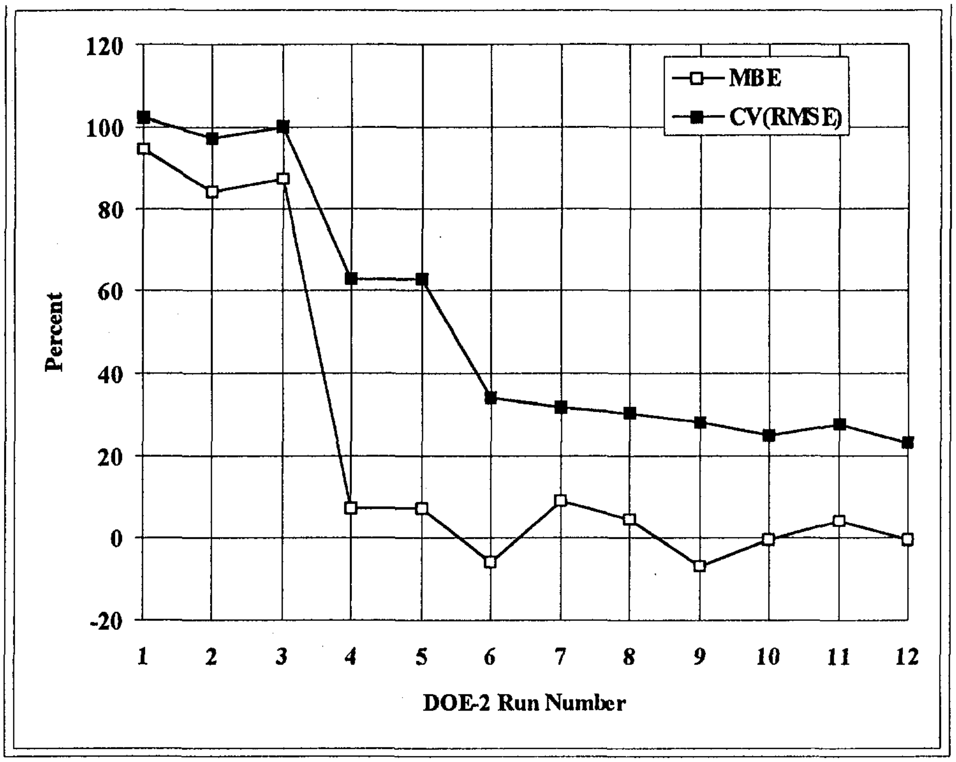

Fig. 9 Tuning progress with input modification. This figure illustrates the progress of the tuning process detailed in Table 2

Conclusions

This paper has investigated techniques for improving calibrated computer building energy simulation methods and has presented several new techniques for improving calibrations. The new methods include graphical procedures and statistical goodness-of-fit parameters for quantitatively comparing simulated data to measured data. A four zone, single story electrically heated and cooled case study building was simulated with DOE-2.1D and calibrated using hourly measured whole-building electricity data and ambient weather conditions to demonstrate the new techniques.

Findings from Applying the New Techniques. The important new calibration features that were developed (Bou-Saada, 1994a) include:

- The use of architectural rendering software allows for visual verification of size and placement of the building's exterior surfaces and shading surfaces to those of the actual building.
- The combined use of a statistical analysis and data display improves the assessment of weather-independent and weather-dependent calibrations.
- The use of a DOE-2 zone temperature comparison to measured indoor temperature confirms simulated HVAC schedules.
- The use of solar beam and diffuse data synthesized from on-site measured global solar radiation data had a modest improvement on this simulation.[5]

These new techniques are a significant improvement over the previously reported DOE-2 calibration methods. The long-term goal of this type of research is to eventually lead to a standardized calibration procedure that could be used on a wide variety of buildings and simulation codes. The next section provides additional details concerning recommendations for future work.

Recommendations and Calibration Procedure. During this research effort, many valuable lessons were learned by refining the simulation. As a guide to future DOE-2 users, recommendations are provided so that calibration efforts can be improved. The first step to simulating a building should include a site visit to acquire the following building information:

- A complete set of as-built drawings (architectural, mechanical, and electrical). Information gathered from as-built drawings includes envelope description and placement, HVAC zoning, lighting loads, and control specifications.
- An HVAC air balance report including supply and return air temperature measurements and air flows.
- Information concerning actual thermostat settings including day/night setback. If an EMCS is installed, a print of the program settings is helpful.
- Measured indoor temperature in each zone during normal operations and nighttime for heating, cooling, and intermediate seasons.
- Hourly HVAC schedules, hourly interior/exterior lighting schedules, hourly occupancy schedules, and hourly equipment schedules. Note the general operation schedule for miscellaneous activity such as dishwashing frequency, etc.
- Perform blower door test to check infiltration rates.
- Evaluate exterior shading surrounding the building including relative distance from building off-south orientation, height, and angle of shading objects.
- Measure at least seven to nine months of hourly whole-building electricity, cooling, heating, and major equipment end-uses.

[5] The use of measured beam and diffuse solar radiation can have a significant impact on those sites that have a strong solar influence.

Table 2 DOE-2 Input file changes for major groups of runs. This table lists the BDL changes that were made between the groups of runs and indicates the best MBE and CV(RMSE) that were achieved for each run.

Run	Change Made to Model	MBE (%)	CV(RMSE) (%)
1	Base Model	94.8	102.4
2	Adjusted infiltration from 0 to 0.1 ACH Added PLANT sub-program to model for hrly reports Modified indoor light & indoor equipment schedules from 10% night/100% day to 35% night/100% day	84.0	97.4
3	Set correct cooling & heating capacities (as per spec.)	87.6	100.0
4	Modified indoor light & indoor equipment schedules (adjusted schedules downward)	7.2	62.8
5	Corrected wall as per DrawBDL Added sizing option-> "ADJUST-LOADS"	7.1	62.8
6	Adjusted air changes from 0.1 ACH to 0.6 ACH Set exhaust CFM as per balance report Set exhaust kW as per balance report Adjusted lighting kW in 1 zone Added surrounding building shading to input file Added ground temperatures to input file Adjusted cooling and heating schedules from 24-hour operation to off at night	-5.9	33.7
7	Adjusted indoor equipment schedules	8.9	31.7
8	Set thermostat settings as per EMCS specs. (summer: 74-day, 80-night; winter: 72-day, 55-night) Added baseboard heat (7504 BTUH) in north zone	4.6	30.1
9	Adjusted indoor light & indoor equipment schedules Modified HVAC schedule to winter/summer setting Added DHW schedule Adjusted fan schedule from 24-hour operation to off at night	-6.9	28.2
10	Removed two buildings from building shading (they were not close enough to affect shading) Adjusted indoor light & indoor equipment schedules Adjusted DHW schedule	-0.4	25.0
11	Adjusted indoor light & indoor equipment schedules Increased return air CFM as per balance report (this corrected an error of not simulating enough fans)	4.1	27.4
12	Added return air fan as per balance report Adjusted indoor light & indoor equipment schedules	-0.7	23.1

- For smaller electrical equipment such as appliances and small motors, clamp-on RMS Watt measurements for 24-hours should be made.
- Survey equipment and note all specifications.
- An accurate light fixture count should include Wattage and ballast power level. Note how many lamps are typically off for extended periods and verify power with clamp-on RMS Watt meter.
- Measure local hourly weather data including relative humidity, dry bulb temperature, wind speed, and global horizontal solar radiation. An alternative can include the NWS, preferably with a weather station as close to the site as is possible. However, NWS solar data are not very useful for simulation purposes, and NWS wind data are peak 5-minute gusts which have almost no correlation to average hourly wind speed.
- Photograph the building's exterior, interior, equipment, and surrounding area to compare with computerized rendering.
- Contact HVAC and internal equipment representative for off-peak equipment performance specifications.

Then create the DOE-2 simulation input file by matching the building HVAC system to the nearest fixed schematic HVAC system in the DOE-2 reference manual. It is imperative to use as many measured details as possible. All DOE-2 default values should be investigated for appropriateness.

Use data from the DOE-2 hourly reports corresponding to the measured data from the building. Process the data into a single ASCII columnar datafile compatible with the comparative graphical routines presented in this paper. Finally, use tools similar to those described in this paper (Bou-Saada, 1994a) and those previously developed (Bronson, 1992; Hinchey, 1991) to compare the simulated energy use to the measured energy use. Iterate until the measured data matches the simulated data to a suitable level as evaluated with hourly MBE, RMSE, and CV(RMSE).

Acknowledgements

Work discussed in this paper was sponsored by the United States Department of Energy through Sandia National Laboratory under contract number AK-8010 and through Oak Ridge National Laboratory under contract number DE-AC05-84OR41200. Certain of the routines also used DOE-2 calibration techniques developed by Bronson (1992) which were supported by the Texas State Energy Conservation Office through the LoanSTAR program and by the Texas ERAP project # 227. We would also like to acknowledge the help provided by all the staff members and the graduate students at the Energy System Laboratory. Special thanks are also due to David Claridge, Robert Sparks, Ron Chambers and Anna Baronwaski for their help over the years.

References

Abbas, M., 1993, "Development of Graphical Indices for Building Energy Data," M.S. Thesis, Energy Systems Rpt No. ESL-TH-93/12-02, Texas A&M University, College Station, TX.

ASHRAE, 1991, "HVAC Applications Volume," *American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc.,* Atlanta, GA.

Ayers, J. M., and Stamper, E., 1995, "Historical Development of Building Energy Calculations," *ASHRAE Transactions Preprint.* Vol. 101, No. 1.

Bou-Saada, T. E., 1994a, "An Improved Procedure for Developing a Calibrated Hourly Simulation Model of an Electrically Heated and Cooled Commercial Building," M.S. Thesis, Energy Systems Rpt No. ESL-TH-94/12-01, Texas A&M University, College Station, TX.

Bou-Saada, T. E., 1994b, "Advanced DOE-2 Calibration Procedures: A Technical Reference Manual," Energy Systems Laboratory Rpt No. ESL-TR-94/12-01, Texas A&M University, College Station, TX.

Bou-Saada, T., and Haberl, J., 1995a, "A Weather-Daytyping Procedure for Disaggregating Hourly End-use Loads in an Electrically Heated and Cooled Building from Whole-building Hourly Data," *Proceedings of the 30th Intersociety Energy Conversion Engineering Conference,* July 31–August 4, 1995, Orlando, Florida, pp. 349–356.

Bou-Saada, T., and Haberl, J., 1995b, "An Improved Procedure for Developing Calibrated Hourly Simulation Models," *Proceedings of the International Building Performance Simulation Association,* August 14–16, 1995, Madison, Wisconsin.

Bronson, J. D., 1992, "Calibrating DOE-2 to Weather and Non-Weather Dependent Loads for a Commercial Building," M.S. Thesis, Energy Systems Rpt No. ESL-TH-92/04-01, Texas A&M University, College Station, TX.

Bronson, D. J., Hinchey, S. B., Haberl, J. S., and O'Neal, D. L., 1992, "A Procedure for Calibrating the DOE-2 Simulation Program to Non-Weather-Dependent Measured Loads," *ASHRAE Transactions,* Vol. 98, No. 1, pp. 636–652. Energy Systems Lab Rpt. No. ESL-PA-91/05-01, Texas A&M University, College Station, TX.

Clarke, J., Strachan, P., Pernot, C., 1993, "An Approach to the Calibration of Building Energy Simulation Models," *ASHRAE Transactions,* Vol. 99, Pt. 2, pp. 917–927.

Cleveland, W. S., 1985, *The Elements of Graphing Data,* Wadsworth & Brooks/Cole, Pacific Grove, CA.

Degelman, L., 1995, "ENER-WIN: Hourly Energy Simulation Model for Calculating Energy Performance in Buildings," Department of Architecture, Texas A&M University, College Station, Texas 77843.

Diamond, S. C., and Hunn, B. D., 1981, "Comparison of DOE-2 Computer Program Simulations to Metered Data for Seven Commercial Buildings," *ASHRAE Transactions,* Vol. 87, No. 1, pp. 1222–1231.

Draper, N., and Smith, H., 1981, *Applied Regression Analysis,* 2nd. Ed. John Wiley & Sons, NY, NY.

Erbs, D., Klein, S., Duffie, J., 1982, "Estimation of Diffuse Radiation Fraction for Hourly, Daily, and Monthly-Average Global Radiation," ASME JOURNAL OF SOLAR ENERGY, Vol. 28, No. 4, pp. 293–302.

Fels, M., 1986, "Special Issue Devoted to Measuring Energy Savings, The Princeton Scorekeeping Method (PRISM)," *Energy and Buildings,* Vol. 9, Nos. 1 and 2.

Grifiths, D. M., and Anderson, K. J., 1994, "Commercial Building/System Characteristics Sensitivity Analysis," *Proceedings from the ACEEE 1994 Summer Study on Energy Efficiency in Buildings,* Vol. 5, pp. 5.105–5.112. Washington, DC.

Haberl, J. S., Bronson, J. D., and O'Neal, D. L., 1995, "An Evaluation of the Impact of Using Measured Weather Data Versus TMY Weather Data in a DOE-2 Simulation of an Existing Building in Central Texas," *ASHRAE Transactions,* Vol. 101, No. 1. ESL Rpt No. ESL-TR-93/09-02. College Station, TX.

Haberl, J., and Claridge, D., 1985, "Retrofit Energy Studies of a Recreation Center," *ASHRAE Transactions,* Vol. 91, Part 2b, pp. 1421–1433.

Haberl, J., and Thamilseran, S., 1996, "Predicting Hourly Building Energy Use: The Great Energy Predictor Shootout II, Measuring Retrofit Savings—Overview and Discussion of Results," *ASHRAE Transactions,* June.

Hadley, D. L., 1993, "Daily Variations in HVAC System Electrical Energy Consumption in Response to Different Weather Conditions," *Energy and Buildings,* No. 19, pp. 235–247.

Hinchey, S. B., 1991, "Influence of Thermal Zone Assumptions on DOE-2 Energy Use Estimations of a Commercial Building," M.S. Thesis, Energy Systems Rpt No. ESL-TH-91/09-06, Texas A&M University, College Station, TX.

Hirsch, J., Winkelmann, F., Buhl, F., Ellington, K., McMenamin, J., Rohmund, I., Criswell, S., Bhimani, A., Spurlock, B., Borstein, D., and Johnson, K., 1995, "PowerDOE, A Windows-based Visually Oriented Analysis Tool," *Proceedings of the 1995 IBPSA Building Simulation Conference,* Madison, WI, pp. 669–673, (August).

Hsieh, E. S., 1988, "Calibrated Computer Models of Commercial Buildings and Their Role in Building Design and Operation," M.S. Thesis, PU/CEES Rpt No. 230, Princeton University, Princeton, NJ.

Huang & Associates, 1993, "DrawBDL User's Guide," 6720 Potrero Ave., El Cerrito, California, 94530.

Hunn, B. D., Banks, J. A., and Reddy, S. N., 1992, "Energy Analysis of the Texas Capitol Restoration," *The DOE-2 User News,* Vol. 13, No. 4, pp. 2–10.

Kaplan, M. B., Caner, P., and Vincent, G. W., 1992, "Guidelines for Energy Simulation of Commercial Buildings," *Proceedings from the ACEEE 1992 Summer Study on Energy Efficiency in Buildings,* Vol. 1, pp. 1.137–1.147.

Kaplan, M. B., Jones, B., and Jansen, J., 1990, "DOE-2.1C Model Calibration with Monitored End-Use Data," *Proceedings from the ACEEE 1990 Summer Study on Energy Efficiency in Buildings,* Vol. 10, pp. 10.115–10.125.

Katipamula, S., 1994, "Personal Communication," Richland, WA, Pacific Northwest Laboratory.

Kreider, J., and Haberl, J., 1994a, "Predicting Hourly Building Energy Usage: The Great Energy Predictor Shootout: Overview and Discussion of Results," *ASHRAE Transactions Technical Paper,* Vol. 100, Pt. 2, June.

Kreider, J., and Haberl, J., 1994b, "Predicting Hourly Building Energy Usage: The Results of the 1993 Great Energy Predictor Shootout Identify the Most Accurate Method for Making Hourly Energy Use Predictions," *ASHRAE Journal,* pp. 72–81, March.

DOE-2 User Guide, Ver. 2.1, 1980, Lawrence Berkeley Laboratory and Los Alamos National Laboratory, Rpt No. LBL-8689 Rev. 2; DOE-2 User Coordination Office, LBL, Berkeley, CA.

DOE-2 Engineers Manual, Ver. 2.1A, 1981, Lawrence Berkeley Laboratory and Los Alamos National Laboratory, Rpt No. LBL-11353; DOE-2 User Coordination Office, LBL, Berkeley, CA.

DOE-2.1 Reference Manual Rev. 2.1A, 1982, Lawrence Berkeley Laboratory and Los Alamos National Laboratory, Rpt No. LBL-8706 Rev. 2; DOE-2 User Coordination Office, LBL, Berkeley, CA.

DOE-2 Supplement, Ver 2.1D, 1989, Lawrence Berkeley Laboratory, Rpt No. LBL-8706 Rev. 5 Supplement. DOE-2 User Coordination Office, LBL, Berkeley, CA.

Mackay, D. J., 1994, "Bayesian Nonlinear Modeling for the Energy Prediction Competition," *ASHRAE Transactions,* Vol. 100, No. 2, pp. 1053–1062.

Manke, J. M., and Hittle, D., 1996, "Calibrating Building Energy Analysis Models Using Short Term Test Data," *Proceedings of the ASME/JSME Solar Energy Conference,* November, 1995.

McLain, H. A., Leigh, S. B., and MacDonald, J. M., 1993, "Analysis of Savings Due to Multiple Energy Retrofits in a Large Office Building," Oak Ridge National Laboratory, Rpt No. ORNL/CON-363, Oak Ridge, TN.

Reddy, S. N., 1993, "Determination of Retrofit Energy Savings Using a Calibrated Hour-By-Hour Building Energy Simulation Model," M.S. Thesis, The University of Texas at Austin. Austin, TX.

SAS Institute, Inc., 1989, "SAS Reference Manual," SAS Institute, Inc., SAS Circle, P.O. Box 8000, Cary, NC, 27512-8000.

Torres-Nunci, N., 1989, "Simulation Modeling Energy Consumption in the Zachry Engineering Center," M.S. Project Report, Texas A&M University, College Station, TX.

Tukey, J. W., 1977, *Exploratory Data Analysis,* Addison-Wesley Publishing Co. Reading, MA.

APPENDIX

The Coefficient of Variation of the Root Mean Square Error (CV-RMSE) and Mean Bias Error (MBE) are described here. These were the statistical indicators that were used in the Predictor Shootout I (Kreider and Haberl, 1994a, b) with the exception of parameter "p" in the definition of CV (RMSE) and MBE which indicates the total number of regression parameters in the model. For the purpose of this evaluation this parameter was assigned an arbitrary value of 1. The definitions of these two indicators are given below.

Coefficient of Variation CV (%):

$$\mathrm{CV(RMSE)} = \frac{\sqrt{\dfrac{\sum_{i=1}^{n} (y_{\mathrm{pred},i} - y_{\mathrm{data},i})^2}{n-p}}}{\bar{y}_{\mathrm{data}}} \times 100$$

Mean Bias Error, MBE (%):

$$\mathrm{MBE} = \frac{\dfrac{\sum_{i=1}^{n} (y_{\mathrm{pred},i} - y_{\mathrm{data},i})}{n-p}}{\bar{y}_{\mathrm{data}}} \times 100$$

where

$y_{\mathrm{data},i}$ is a data value of the dependent variable corresponding to a particular set of the independent variables,

$y_{\mathrm{pred},i}$ is a predicted dependent variable value for the same set of independent variables above,

$\bar{y}_{\mathrm{data}}$ is the mean value of the dependent variable of the data set,

n is the number of data points in the data set, and

p is the total number of regression parameters in the model (which was arbitrarily assigned as 1 for all models).